



## Evaluation de la qualité des vues 3D synthétisées

Emilie Bosc, Muriel Pressigout, Luce Morin

### ► To cite this version:

Emilie Bosc, Muriel Pressigout, Luce Morin. Evaluation de la qualité des vues 3D synthétisées. ORASIS, Jun 2011, Praz-sur-Arly, France. pp.39. hal-00628075

**HAL Id: hal-00628075**

**<https://hal.science/hal-00628075>**

Submitted on 6 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluation de la qualité des vues 3D synthétisées

Emilie Bosc

Muriel Pressigout

Luce Morin

IETR UMR CNRS 6164, Groupe Image et Télédétections/INSA de Rennes  
20, avenue des Buttes de Coesmes, 35708 RENNES CEDEX 7, France  
(prénom.nom@insa-rennes.fr)

## Résumé

*Les données multivues plus profondeur (MVD) nécessitent de méthodes de compression efficaces, compte tenu de leur taille importante. Pour valider ces méthodes de compression, des protocoles d'évaluation sont indispensables. En effet, le défi récurrent en compression reste l'optimisation du rapport entre le débit et la distorsion des données transmises. Dans ce contexte, la métrique objective la plus utilisée est le PSNR (Peak-Signal-to-Noise-Ratio). Cet article se propose d'analyser dans quelle mesure le PSNR est fiable pour l'évaluation de la qualité des vues synthétisées. Pour ce faire, nous réalisons l'encodage des cartes de profondeur uniquement et analysons la qualité des images synthétisées à partir de ces données dégradées, en termes de PSNR. Les codecs utilisés sont H.264 et la méthode LAR (Locally Adaptive Resolution). Les résultats montrent que pour de faibles valeurs de PSNR, il est difficile de conclure sur les performances d'un codec, sans une analyse visuelle supplémentaire.*

## Mots Clef

Compression, vidéo 3D, MVD, profondeur, codage adaptatif.

## Abstract

*Multiview plus depth (MVD) data consist of large amount of data and require efficient compression method. Proposed compression methods performances need to be evaluated through reliable assessment protocols. While the challenge in compression is the optimization of the rate-distortion ratio, a widely used objective metric to evaluate the distortion is the Peak-Signal-to-Noise-Ratio (PSNR), because of its simplicity and mathematical easiness to deal with such purposes. This paper addresses the reliability, concerning this metric, in the multiview video context. We investigated the visual performances of two methods, namely H.264/MVC and Locally Adaptive Resolution (LAR) method, by encoding depth maps and reconstructing existing views from those degraded depth images. Results show that when using a pixel-based metric such as PSNR for assessing virtual synthesized views, low scores do not imply low coding performances.*

## Keywords

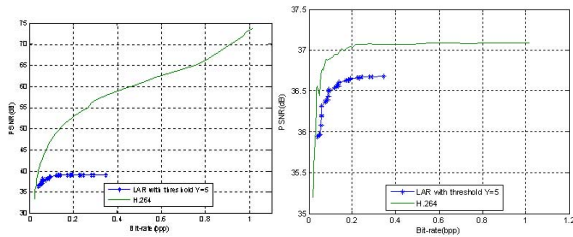
3D video coding, depth coding, adaptive coding.

## 1 Introduction

Après la révolution de la télévision couleur dans les années 40, le nouveau bouleversement attendu est la vidéo tridimensionnelle. Elle suscite l'intérêt grâce à ses deux applications possibles : la télévision tridimensionnelle, ou TV 3D, qui donne une impression de profondeur ou de relief à la scène ; et le libre choix de point de vue (*Free Viewpoint Television*, FTV) qui permet de naviguer interactivement dans la scène. Ces applications sont rendues possibles grâce au format particulier des données MVD [1]. Elles sont constituées d'un ensemble de vidéos couleur conventionnelles, acquises à des points de vue légèrement différents de la même scène, ainsi que d'un ensemble de vidéos en niveaux de gris contenant les informations de profondeur associées. Chaque pixel de la vidéo de profondeur indique la profondeur du pixel de couleur correspondant, c'est à dire sa coordonnée Z dans le repère caméra.

Compte tenu de la quantité de données à traiter, des méthodes de compression efficaces sont requises. Il n'existe pas de standard de compression pour les données MVD. En effet la difficulté de conception d'une méthode robuste réside essentiellement dans le fait que l'impact des dégradations de compression sur les données de profondeur peut être fatal à la qualité de la vue synthétisée, comme cela a été montré dans de récentes études [2, 3]. En effet, des algorithmes appelés Depth Image Based Rendering (DIBR) [4] permettent de reconstruire des vues virtuelles, c'est-à-dire non acquises par les caméras, à partir des données de couleur et de profondeur des vues adjacentes. Des artefacts liés aux méthodes de compression utilisées peuvent induire des erreurs de projection perceptibles sur la synthèse. La plupart des études se basent sur la mesure du PSNR comme argument de validation des performances, car il est la mesure de référence pour l'évaluation de méthodes de compression pour des vidéos conventionnelles. Il s'exprime comme suit :

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{EQM},$$
avec  $EQM = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I_r(i, j) - I_v(i, j))^2$   
où  $I_r$  est l'image de référence,  $I_v$  la vue rendue,  $M \times N$



(a) Distorsion de la profondeur (b) Distorsion la vue synthétisée  
FIGURE 1 – Courbes de débit/distorsion pour "Breakdancers", avec LAR et H.264.

sont les dimensions de l'image. Cependant, cette mesure, basée pixel, ne traduit pas les nouveaux artéfacts liés au processus même de synthèse de vue. En effet, il peut arriver qu'une zone entière de l'image, correspondant à un objet cohérent sémantiquement, soit translatée lors de la projection, à cause de valeurs de profondeur approchées. Cela entraîne nécessairement une chute du score du PSNR, bien que l'image rendue reste de bonne qualité du point de vue perceptuel. A l'inverse, des erreurs localisées sur les discontinuités des cartes de profondeur, provoquent des erreurs de projections concernant peu de pixels : on peut donc obtenir un PSNR correct alors que le résultat est gênant du point de vue perceptuel.

Les expériences réalisées cherchent à déterminer l'impact de la compression des données de profondeur sur la qualité de la vue rendue. Elles visent également à déterminer si les résultats objectifs d'évaluation des performances de codage (PSNR) sont en accord avec l'appréciation visuelle.

## 2 Protocole et Résultats

Deux codecs ont été utilisés et comparés : le codec standard H264 basé blocs et DCT[5], et le codec LAR conçu pour la préservation de la qualité visuelle et basé sur une décomposition aplats/textures et un quad-tree adaptatif [6]. Seul l'encodage des cartes de profondeur est réalisé, les images de couleur restant originales : deux points de vue sont considérés, la vue 2 et la vue 4 de la séquence *Breakdancers*. Après le décodage, la synthèse du point de vue intermédiaire (vue 3) est obtenue grâce à l'algorithme VSRS (View Synthesis Reference Software, version 3.5), fourni par MPEG. Le PSNR des cartes de profondeur décodées ainsi que des images de synthèse générées est donné par la figure 1. Elle montre que les performances en termes de PSNR de la méthode H.264 dépassent celles de la méthode LAR. Notons l'écart important pour la distorsion de la profondeur. En revanche, l'écart observé pour la distorsion de la vue synthétisée est plus réduit (0.5dB). Cela indique que la méthode LAR conserve correctement des données essentielles à la synthèse. La figure 2 montre des zones de la vue synthétisée pour 0.04bpp. Pour ces images, bien que le PSNR de la méthode LAR soit moins élevé que celui de la méthode H.264, on constate que la qualité des vues rendues est proche ou légèrement améliorée avec la méthode LAR : la ligne verticale ainsi que le contour du visage sont mieux préservés.

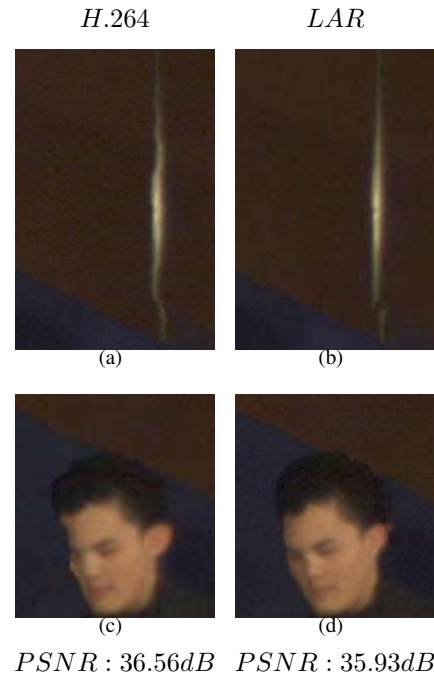


FIGURE 2 – Images rendues avec VSRS (profondeur encodée à 0.04bpp).

## 3 Conclusion

Nous avons étudié l'impact de la compression des cartes de profondeur sur la qualité de la vue synthétisée à partir de deux points de vue et avec deux codecs. Les résultats montrent que la compression H.264 obtient un meilleur PSNR. Cependant, l'analyse visuelle des images rendues permet de constater qu'une méthode comme le LAR, obtenant un plus faible PSNR, permet de rendre des vues d'une qualité visuelle similaire voire meilleure. Ainsi, cela indique que des méthodes plus orientées sur le respect de la qualité perçue peuvent être de bonnes candidates pour la compression des données MVD, et que le PSNR seul est inadapté pour l'évaluation des performances.

## 4 Remerciements

Ces travaux ont été réalisés dans le cadre des projets ANR-PERSEE (ANR-09-BLAN-0170), ANR-CAIMAN (ANR-08-VERS-002). Nous souhaitons également remercier Microsoft Research pour la mise à disposition de la séquence multi-vues "Breakdancers", et MPEG l'algorithme de synthèse.

## Références

- [1] A. Smolic, K. Mueller, P. Merkle, N. Atzpadin, C. Fehn, M. Mueller, O. Schreer, R. Tanger, P. Kauff, and T. Wiegand. Multi-view video plus depth (MVD) format for advanced 3D video systems. *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q. 6, JVT-W*, pages 21–27, 2007.
- [2] Emilie Bosc, Vincent Jantet, Luce Morin, M. Pressigout, and Christine Guillemot. Vidéo 3D : quel débit pour la profondeur ? In *Proc. CORESA*, Lyon, 2010.
- [3] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, and T. Wiegand. The effects of multiview depth video compression on multiview rendering. *Singal Processing : Image Communication*, 24(1-2):73–88, 2009.
- [4] C. Fehn. 3D-TV using depth-image-based rendering (DIBR). In *Proc. (PCS)*, pages 307–312, 2004.
- [5] ITU-T Recommendation H.264. Advanced video coding for generic Audio-Visual services, 2009.
- [6] O. Deforges, M. Babel, L. Bedat, and J. Ronsin. Color LAR codec : a color image representation and compression scheme based on local resolution adjustment and self-extracting region representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(8):974–987, 2007.